知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

印本馆藏大规模数字化的研究与探讨

◎ 李咏梅[1] 刘柳[1] 刘鸿[2]
[1] 四川大学图书馆 成都 610065 [2] 四川建筑职业技术学院图书馆 德阳 618000

摘要：[目的 / 意义] 印本馆藏大规模数字化是数字图书馆发展的必然之举，它既无限地扩大了图书馆服务的深度和广度，同时又极大地推动了学术交流和学术研究的发展进步。[方法 / 过程] 从发展缘起、环境因素、发展现状和存在的问题等方面对全球印本馆藏大规模数字化运动进行分析研究，并以美国 HathiTrust 项目作为实证分析案列。[结果 / 结论] 确立大规模数字化原则、立足合理使用原则解决版权问题、建立多级多维度合作框架是促进印本馆藏大规模数字化长足发展的重要举措。

关键词：印本馆藏 大规模数字化 云图书馆 HathiTrust 合理使用
分类号：G253

引用格式：李咏梅 , 刘柳 , 刘鸿 . 印本馆藏大规模数字化的研究与探讨 [J/OL]. 知识管理论坛 , 2017, 2(5): 406-413[引用日期]. http://www.kmf.ac.cn/p/1/657/.

始于 2004 年的 Google 图书计划，掀起了印本馆藏大规模数字化运动的浪潮。2008 年 HathiTrust 宣告成立，把这一运动推向了高潮。2014 年 2 月，HathiTrust 数字图书馆里的藏书突破了 1 100 万册 [1]，迎来了 HathiTrust 发展史上的一个里程碑；2017 年 2 月，HathiTrust 的馆藏突破了 1 500 万册 [2]，实现了新的飞跃。星星之火已经燎原，全球图书馆开展的数字化运动极大地推动了学术交流和学术研究的发展进步。

1 缘起

传统的印本资源通常要经过出版、传播和获取三个环节后才能到达最终用户手上，并通过最终用户的使用来实现其价值。对于任何一个想要阅读或者使用资源的人来说，总是希望能更快、更便捷、不受时空限制、完全无障碍地获取到其所需，这也是图书馆一直所追寻的终极目标。如果最终用户的数量达到极致，也就是说，印本资源被尽可能多的人使用，其价值将得到最大程度的彰显。而印本资源的大规模数字化正是实现这一目标的重要手段，它使得印本资源的使用价值最大化，极大地推动了人类文明和知识的传播，并促进了学术交流和学术研究的深度和广度发展。

2004 年 12 月，Google 宣告启动 Google Print 图书馆计划，与哈佛大学、密西根大学、纽

基金项目：本文系四川大学“图书馆、情报与文献学”科研项目“数字资源对图书馆文献资源建设的影响研究”(项目编号：sktq201109）研究成果之一。
作者简介：李咏梅 (ORCID：0000-0001-6891-9871)，采访中心主任，研究馆员，硕士，E-mail：yongmeili@scu.edu.cn；刘柳 (ORCID：0000-0002-4643-4512)，采访中心副主任，副研究馆员；刘鸿 (ORCID：0000-0002-5745-9088)，副馆长，副研究馆员。
收稿日期：2017-07-21 发表日期：2017-10-13 本文责任编辑：王传清

知识管理论坛，2017(5):406-413
DOI：10.13266/j.issn.2095-5472.2017.044

约公共图书馆、牛津大学和斯坦福大学等几所超大型图书馆进行合作，这些馆馆藏量估计超过 1 500 万册 [3]。双方达成协议：由 Google 出资提供技术设备和对馆藏资源进行扫描和编目，使全世界的读者可从 Google 网站上搜索读取资源全文的片段、摘要及其馆藏和市场信息。几年后，谷歌已经扫描了近千万册书籍供网上搜索 [4]。

2008 年 4 月，荷兰图书馆首开先河，与版权所有者签署意向宣言，就遗产收集的数字化和可获取性达成共识，这是全球第一份图书馆与版权所有者签订的此种类型的协议，其核心内容在于图书馆可得到版权所有者的授权，数字化其作品并在用于教学或科研目的前提下在本馆馆舍内供公共获取 [5]。这有力地推动了传统馆藏大规模数字化运动的开展。

2008 年 10 月，HathiTrust 宣告诞生，它是由美国机构合作委员会（CIC）图书馆联盟和加州大学图书馆联盟共同创建的共享数字仓储项目，将其成员馆所收藏的印刷型文献进行数字化，并统一保存 [6]。至此，印本馆藏的大规模数字化运动蓬勃开展起来。

2 大规模数字化的环境因素

2.1 数字图书馆的大发展

随着数字化和网络化的发展，越来越多的资源以网络为载体，网络数字资源跨越时空的方便获取性，使得在线阅读成为新型阅读方式，并日益成为主流。用户不断变化的阅读习惯和行为模式使得图书馆必须一直在新技术和用户体验方面保持领先优势，这是数字图书馆发展的原动力，对数字馆藏的需求增加以及移动服务的新需求，使得传统纸本馆藏很难再满足用户在线阅读需求，大规模数字化成为数字图书馆发展的重要组成部分，实现图书馆最大化提供馆藏和服务的在线利用。同时大规模数字化客观上减少了对空间和人力的需求，从而降低了馆藏管理成本。

此外，随着移动技术的日新月异，移动设备正改变着人们传递和获取信息的方式，移动阅读日渐兴起，由于其具有阅读工作的便携性和可移动性、阅读内容的可检索性和及时获取性、阅读行为的持久性和连续性等优势 [7]，逐渐成为一种主流阅读模式，这对数字图书馆服务内容的多样性和丰富性提出了更高的要求。为适应移动环境的发展变化，图书馆开始利用移动设备来提供服务和内容传递，用户可以借助手机来方便地利用数字化后的印本资源。而大数据时代的兴起又对图书馆的数字化发展提出了更迫切的要求。

2.2 开放获取运动的大发展

开放获取运动的理念就在于要实现研究成果能被任何想要阅读和使用的人无障碍地任意获取，从而使得科研成果的社会、经济和文化效益得到最大程度地发挥。资源的开放、透明和易于获取已逐渐成为一种价值观念。

2007 年 12 月美国政府通过了 NIH 实行强制性开放存取的法律条款；2012 年美国联邦研究成果公共获取提案 (FRPAA) 再次被提交至众议院和参议院，将强制开发获取的领域扩大到所有联邦机构；欧洲研究型大学联盟也发布了《开放研究数据》和《研究出版物的开放获取》的声明 [8]。

作为全球最大的科研资助机构之一的英国惠康基金会一直是开放获取的积极倡导者和领跑者，早在 2006 年就要求其资助的原创研究论文在正式出版后 6 个月内存储到 Europe PubMed Central 中开放获取，2012 年推出了更为强硬的存储政策，要求 2013 年 10 月起资助的新项目以及 2014 年 10 月起所有被资助项目所出版的专著或图书章节实行与论文相同的开放获取政策 [9]。强制开放获取政策极大地推动了开放获取的发展。

开放获取运动对学术研究领域所产生的影响是前所未有的，它促进了学术成果数字化的发展，同时数字化的发展又进一步推动开放获取运动的快速进步，为全世界的用户提供了更为广泛而深入的文献资源的公开获取，而印本



馆藏的数字化是图书馆顺应开放获取环境下加强自身核心竞争力的必然选择。

2.3 高等教育的发展趋势

高等教育正面临前所未有的机遇和挑战，进行着重大的变革，传统的教育体制和模式日益受到新技术发展的挑战，高等教育机构之间的竞争也愈发激烈，并日趋向国际化方向快速发展。欧美国家的大学纷纷在亚洲和中东地区建立分校，如纽约大学在中国和阿联酋建立分校，诺丁汉大学在中国和马来西亚建立分校，耶鲁大学在新加坡建立分校，跨洲合作办学和授予学位的机构更是层出不穷。

与此同时，由于网络和平板电脑的大量应用，在线教学和学习普及起来，而大规模开放网络课程（MOOC）异军突起，凭借其大量免费内容和高水平的教学越来越受到欢迎，并作为传统大学课程的替代和补充逐渐获得广泛应用 [10]。

高等教育和学术研究的变革无疑会对图书馆的以下几个方面产生重大影响：馆藏的建设和发展、资源的获取和保存以及不断扩展用户的服务等 [11]。把印本馆藏包括各种特色和珍稀馆藏数字化，可以为教学和科研提供更为丰富的网络资源，以满足其所属机构在线教学和学术创新的需要，增强竞争力。

3 发展现状及存在的问题

3.1 发展现状

截至到 2012 年 3 月，Google 已对 2 000 多万册图书进行了数字化 [12]，而 HathiTrust 经过 5 年多的发展，其馆藏已达到 11 262 697 册 [13]。图 1 是 HathiTrust 2009 年 6 月至 2014 年 6 月 5 年期间数字馆藏年度统计 [14]，从图 1 中可以看出，5 年期间其馆藏增长了 240%，平均每年增幅接近 50%，增速是很快的。

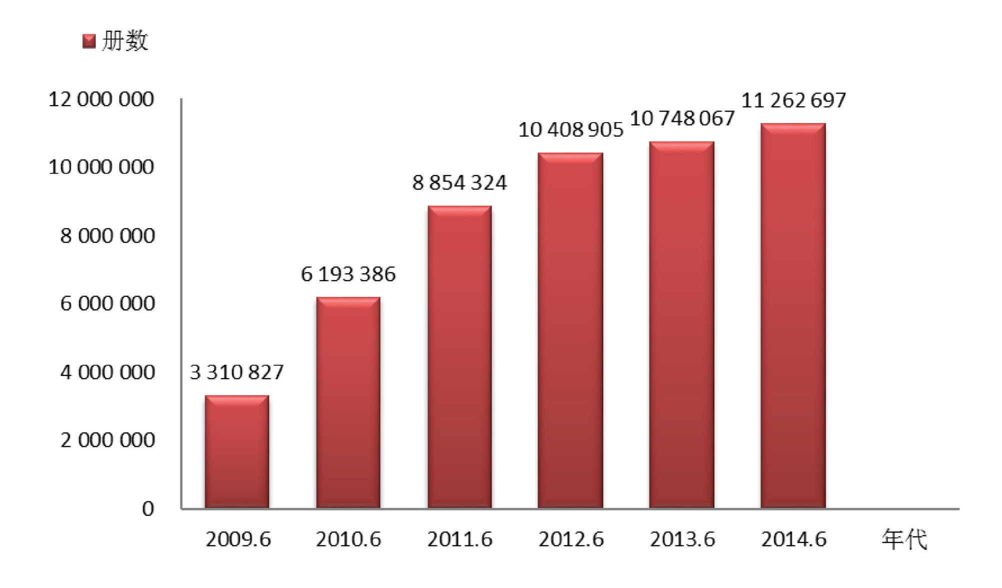


图 1 HathiTrust 2009.6–2014.6 数字馆藏年度统计

从图 1 还能看出，HathiTrust 的馆藏在 2012 年后增长势头减缓，这是由于 Google 放慢了图书扫描工作，这反映了该项目的自然成熟过程。拟扫描的图书存在重复的可能增加，因此采集量减少了，同时有些机构打破与谷歌的常规协议，只扫描特色馆藏，如德克萨斯大学奥斯汀分校只扫描其拉丁美洲馆藏 [12]。

与此同时，HathiTrust 对进入公共领域的馆藏建设也非常重视，馆藏量逐年不断增长，在总馆藏中的比重越来越大，这点从表 1“HathiTrust 2009.6-2014.6 进入公共领域的馆藏统计”[14] 可以看出，截至到 2014 年 6 月，有 34% 的馆藏可以

知识管理论坛，2017(5):406-413
DOI：10.13266/j.issn.2095-5472.2017.044

免费获取全文，且是面向全球开放的。到2017年2月，有5 811 123册进入公共领域，占15 101 537册总馆藏的38.5%[15]。

表1 Hathitrust 2009.6-2014.6进入公共领域的馆藏统计(单位：册)

年度	总馆藏	公共领域馆藏	占比
2009.6	3 310 827	534 338	16%
2010.6	6 193 386	1 208 351	20%
2011.6	8 854 324	2 413 921	27%
2012.6	10 408 905	3 105 587	29%
2013.6	10 748 067	3 405 432	31%
2014.6	11 262 697	3 848 472	34%

HathiTrust发布的2014年6月统计报告还显示，密西根大学图书馆、加州大学图书馆扫描的馆藏量位居第一和第二，分别为4 689 072册和3 520 634册[13]，分别占HathiTrust总馆藏的41.6%和31.3%，合计所占比例达到73%，可见这两个图书馆是HathiTrust最主要的贡献者。

HathiTrust很多成员馆都提供数字馆藏打印服务，如加州大学图书馆将其数字馆藏已进入公共领域的图书向读者提供重印本服务，并通过亚马逊提供快递服务。而Google更是推出随用随印服务，它设计出一款快速印书机，读者只要付大约8美元就能在5到10分钟内得到一本装订好的简装本[4]，既经济又高效。

HathiTrust不断发展成员馆，让成员馆把本地数字化馆藏存入HathiTrust里，这也是HathiTrust馆藏增长迅速的原因之一，同时HathiTrust还不断开发新的工具(如SIP)以使各成员馆简化和加速存储过程。

欧洲也开展了大量的数字图书馆建设项目，如大英图书馆实施了大量的数字化项目来将本馆的文化遗产数字化，使大量珍贵的手稿既有了数字化替代馆藏，又能被公众广泛获取利用。该馆近期还参与了一些全球性的大规模数字化项目，包括Google图书项目和brightsolid报纸数字化项目[16]。

我国的数字化建设主要以特色馆藏为主，特别是中文古籍的数字化取得了很大的进展。据统计，2012年我国公藏机构已拥有超过20亿字的数字化文本格式的古籍[17]，同时国际合作项目日益增多，其中国家图书馆承担的最多。高校的国际合作项目最具代表性的是CADAL（大学数字图书馆国际合作计划）。截至2014年2月，CADAL馆藏量为2 748 688册/件，数字化的文献包括古籍、图书、期刊、学位论文等。其中中文图书有1 130 595册（含古籍236 581册），外文图书767 059册[18]，是目前国内图书馆对印本图书进行大规模数字化的代表项目，任何用户通过注册就可以免费阅读和打印。

3.2 存在的问题

印本馆藏大规模数字化在快速发展过程中历经了重重困难，存在的以下问题成为其发展的障碍。

3.2.1 版权问题

Google图书计划从实施一开始就一直备受争议，而所有的争议都和版权有关，大规模数字化的发展史也是一部Google的诉讼史。从2005年美国作家同业协会和出版业协会把谷歌告上法庭开始，Google的侵权官司就一直没有停止过，起诉Google的行业组织和出版商越来越多。2011年3月，美国法院驳回了Google图书“和解协议”[19]，给大规模数字化的发展蒙上了阴影，图书馆版权联盟就此发表声明：“这项裁决清楚地表明，现行版权法仍然对图书馆以及其他有兴趣从事大规模数字化项目的合作者造成重大阻碍”[20]。HathiTrust也受到版权问题困扰，美国作家协会起诉其数字图书馆违反了美国版权法。2013年向公众推出在线免费获取的美国数字公共图书馆(Digital Public Library of America, DPLA)也指出版权问题是其目前所面临的最大问题之一，可见，版权问题始终是印本资源大规模数字化的一大障碍。

3.2.2 长期管理与保存

数据的管理和保存是所有数字化项目最重

要也是最困难的工作。通常数据的存储包括书目数据和全文数据两个部分。数字资源不同于实体资源，如果没有得到特别的保护，其生命周期是很短暂的。随着数字图书馆发展的日益成熟，数字馆藏的数量不断增长，其长期管理和保存的问题越来越突出，特别是个体图书馆在长期保存问题上的认识和规划存在较大的不足，这对本地数字馆藏的可持续性发展极为不利。图书馆在创建新资源上的投入要远远多于维护或优化其已创建的资源，即一旦将馆藏数字化，往往基本上不主动对其进行管理[21]。

Portico 和康奈尔大学图书馆的报告表明，大多数文化遗产机构仅仅开始意识到其在建立数字馆藏方面的大量投资必须能够保证在其生命周期内保护这些内容并提供相应的基础设施[22]。OCLC 对美国和加拿大 169 个研究型图书馆特色馆藏数字化的调研显示，有 97% 的图书馆已经完成一个或多个数字化项目。而缺少资金、专业知识和行政支持被认为是其进行数字馆藏管理与保存的最大障碍[23]。

3.2.3 元数据互操作性问题

元数据是有效组织和揭示数字资源必不可少的元素，由于从事大规模数字化项目的机构各种各样，以及数字资源的内容和形式也是丰富多彩，造成了多种元数据标准并存的局面，以及对于元数据的收集和整理困难重重，使得数字资源的整合利用以满足用户统一检索的需求充满了挑战，这必然制约了数字图书馆的可持续性发展。解决这一问题的关键就是实现元数据的互操作性。有学者对大学数字图书馆国际合作计划（CADAL)、中国国家图书馆数字图书馆工程 (数图工程)、谷歌图书 (Google Book)、欧洲数字图书馆 (Europeana)、开放图书馆 (Open Library)、HaithTrust、加州数字图书馆 (CDL)、美国记忆 (AM) 等国内外知名的大型数字化项目的元数据互操作性进行了调研[24]，从中可以看出取得了一定的进展，但要从根本上解决这一问题，还有很长的路要走。

4 关于印本馆藏大规模数字化发展的思考和建议

大规模数字化历经 10 余年的发展，取得的成就是举世瞩目的，它不仅推动了数字图书馆的发展，同时促进了学术研究和学术交流广度和深度的进一步扩展，特别在对知识的整合与创新上成效显著。但其发展过程中遇到的困难和存在的问题成为制约其发展的障碍，如何克服这些困难解决好存在的问题以更好地推动其发展，笔者有以下一些思考和建议。

4.1 确立大规模数字化原则

美国研究图书馆协会（ARL）在 2010 年 6 月发布了大规模数字化九项原则[25]，作为其成员馆在对特藏资源进行大规模数字化的同时，也对供应商或者出版社进行合作方面的指导，并敦促他们使用这些原则。借鉴 ARL 的指导原则，结合上述大规模数字化的现实发展情况，探讨提出我国图书馆在进行印本馆藏大规模数字化时应遵循的原则。

4.1.1 合理使用原则

合理使用原则是著作权法中协调著作权人和使用者之间利益的一项制度，是对知识产权保护与公众利益的一种折中体现。在数字化时代，知识产权的环境日益复杂化，合理使用原则的利用显得更为重要，但由于人为的主观性影响，要准确把握合理使用的度不是一件容易的事情，图书馆在遵循这一原则时，一定要明确合理使用的使用对象和目的，从而在为公众提供服务的同时，最大限度地保护著作权人的利益。

4.1.2 一次性 (扫描) 原则

为了保护印本和特藏资源，特别是那些珍稀资源，要尽可能遵循一次性扫描原则。数字化扫描对印本资源始终都有不同程度的损害，这需要图书馆进行周密的组织和规划，制定出具有前瞻性的扫描计划。

4.1.3 保留副本及更新原则

图书馆应得到从其馆藏中产生的所有数字化文件的副本并完整保留存储，同时对这些

知识管理论坛，2017(5):406-413
DOI：10.13266/j.issn.2095-5472.2017.044

数字化内容进行定期更新，在版权法允许的范围内向相关用户提供这些副本的无限制本地访问，并行使合理使用的权利。

4.1.4 共享原则

图书馆的天然使命是传承和传播知识和文化，在开放获取的大背景下，共享已演变为图书馆界的"普世"价值。图书馆间的数字馆藏进行共享，能避免其他机构的重复投入，从而避免了资源、资金和人力的浪费，HathiTrust 的共享数字仓储项目就是这一原则的最好实践。而共享需要合作的手段才能实现，这将在下面进行详述。

4.2 立足合理使用原则解决版权问题

版权问题是印本馆藏大规模数字化进程中所面临的最大障碍，图书馆需要在遵循著作权法的前提下科学、合理的使用"合理使用"原则才能解决好版权问题。

美国作家协会曾向南纽约地区法院起诉 HathiTrust 违反了美国版权法，2012 年 10 月法院对此案做出裁决，判定 HathiTrust 数字图书馆对印本资源进行扫描和数字化的工作事实上属于美国版权法规定的合理使用。这是合理使用和美国图书馆的一次重大胜利 [26]。随后，美国作家协会又向美国第二巡回上诉法院提起上诉，该法院于 2014 年 6 月做出驳回上诉并维持原判的判决 [27]。而最令人鼓舞的消息是该上诉法庭于 2015 年 10 月宣判，谷歌对图书进行数字化扫描的行为，完全合法，是对图书的公平、正常的使用，对社会有益 [28]。这一判决宣告了这一长达 10 年诉讼案的终结，宣告了谷歌公司的最终胜利，从而也在一定程度上扫清了大规模数字化的障碍。

由此可见，合理使用这一指导原则在解决版权问题中具有极大的指导意义和现实价值。但由于合理使用对著作权人的利益在客观上造成了一定程度的损害，国际著作权法对合理使用有着严格的限制和判断标准，以防止合理使用的过度使用和滥用。

4.3 建立多级多维度合作框架

合作是图书馆事业通向成功之路的法宝，也是实现资源共享向纵深发展的必然手段。Google 图书馆计划和 HathiTrust 项目本身就是合作的产物，Google 和 HathiTrust 在版权诉讼中取得的最终胜利与美国图书馆版权联盟发挥的积极作用分不开。

由 OCLC、HathiTrust、纽约大学 Elmer Holmes Bobst 图书馆以及研究型馆藏获取和保存联盟（ReCAP）联合实施的项目——云图书馆计划（Cloud Library Project）就是一个很好的实践，这一项目得到安德鲁梅隆基金会的大力资助。这是图书馆之间、图书馆和资助者之间的强强合作，是一个赋有前瞻性和可持续性发展的项目，值得借鉴和推广。该项目做的调研显示，截至到 2010 年 6 月，美国研究图书馆学会（ARL）成员馆馆藏资源有 30%（中位数）与 HathiTrust 数字图书馆重复；纽约大学 Bobst 图书馆的馆藏从 2009 年 9 月到 2010 年 6 月期间与 ReCAP 和 HathiTrust 两个数字仓储都重复的数量逐月上升，到 2010 年 6 月重复量达到约 20 万种 [29]，如果 HathiTrust 能提供完善的服务，重复的印本馆藏可以从该馆撤除，立竿见影的受益就是可以节约不少空间和管理成本。目前 HathiTrust 已拥有 1 500 万册的数字馆藏，是 2010 年时的 2.42 倍，保守估计的话，ARL 成员馆馆藏与 HathiTrust 重复的比率已达 60%（中位数）。

云图书馆计划的目标是使印本馆藏在大规模数字化环境下得到颇为高效的管理和最优化配置和整合，从而促进馆藏建设的重新定位和图书馆的服务创新。尽管 HathiTrust 的数字仓储已拥有 1 500 万册图书，如果要实现全面的印本馆藏的云图书馆计划，还需要与其他大型服务机构进行合作，共享馆藏资源，如美国国会图书馆是印本图书的存储机构，虽然美国国会图书馆早在 2011 年就已是 HathiTrust 的成员馆，但其在 HathiTrust 里的数字化馆藏只有 10 万余册，如果 HathiTrust 能和美国国会图书馆有更全面深入的合作，至少可以覆盖 70% 以上的数字化图书资源。HathiTrusth 和美国国会图书馆都

是美国数字公共图书馆的重要合作伙伴 [30]，三者的深入合作将极大地扩大云图书馆计划的范围和影响。

HathiTrust 于 2016 年正式启动的共享印本计划是在共享数字仓储项目基础上更深入的合作与共享，其目标是通过印本及其数字化馆藏的链接来确保二者的保存，从而减少成员馆馆藏管理的费用，同时促进国家乃至洲际馆藏的共同管理和共享。该项目分两个阶段实施，第一阶段在 2017 年取得实质性进展：在第一阶段作为保留伙伴加入该计划的 50 多个成员馆需在 2017 年 4 月底以前确认其计划保留的印本图书 [31]。共享印本计划不仅使 HathiTrust 里数字化馆藏的印本资源得到安全保留的保障，而且保证了成员馆之间印本馆藏的长期可借性，从而为云图书馆计划的最终实现提供更为可靠的保障。

洲际合作也日益盛行。欧洲数字图书馆和美国数字公共图书馆宣布合作，任何用户将可以获得两个系统联合的丰富馆藏，合作的内容包括：可互操作的数字模型、共享的资源规范以及合作的馆藏建设 [32]。

印本馆藏大规模数字化的最终成功需要研究机构、大学、研究者、图书馆、出版商、版权所有者及资助机构的通力合作，建立一种所有利益方都接受的均衡体制，并使之具有可持续性。

参考文献：

[1] Update on February activities 2014[EB/OL]. [2017-01-25]. http://www.hathitrust.org/documents/hathitrust-update-201402.pdf.

[2] February 2017 update[EB/OL]. [2017-04-06]. https://www.hathitrust.org/february-2017-update.

[3] Google Books history[EB/OL]. [2017-01-02]. http://books.google.com/googlebooks/about/history.html.

[4] 邹际平 , 董小芬 . 谷歌图书馆 , 还是图书馆的谷歌化——美国图书馆资源共享回顾和前瞻 [J]. 图书与情报 , 2010(5): 65-69.

[5] 郭明珠 . 传统馆藏数字化指日可待——图书馆与版权所有者达成协议 [J]. 图书情报工作动态 , 2009(2): 38-39.

[6] 李咏梅 , 袁学良 , 唐李杏 . 美国 HathiTrust 项目及其对我国的启示 [J]. 四川图书馆学报 , 2011(3): 35-37.

[7] 赵乃瑄 , 周静珍 . 移动阅读环境下大学图书馆的转型、创新和合作 [J]. 图书情报工作 , 2012, 56(9)：48-51.

[8] LERU launches statements on open access and open data[EB/OL]. [2017-04-02]. http://www.leru.org/index.php/public/news/leru-launches-satements-on-open-access.

[9] Wellcome Trust extends open access policy to include scholarly monographs and book chapters[EB/OL]. [2017-01-02]. http://www.wellcome.ac.uk/News/Media-office/Press—releases/2013 /WTP052746.html.

[10] The New Media Consortium. Horizon report 2013 higher education edition[EB/OL]. [2017-06-01]. http://www.nmc.org/pdf/2013-horizon-report-HE.pdf.

[11] The ACRL Research Planning and Review Committee. Environmental scan 2013[EB/OL]. [2017-02-06]. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/EnvironmentalScan13.pdf.

[12] HOWARD J. Google begins to scale back its scanning of books from university libraries[EB/OL]. [2017-05-30]. http://chronicle.com/article/Google-Begins-to-Scale-Back/131109/?sid=at&utm_source= at&utm_medium=en.

[13] Update on June 2014 activities[EB/OL]. [2017-01-26]. http://www.hathitrust.org/ updates_june2014.

[14] All newsletters and reports[EB/OL]. [2017-01-26]. http://www.hathitrust.org/updates.

[15] Currently digitized[EB/OL]. [2017-04-06]. https://www.hathitrust.org/visualizations_deposited_volumes_current.

[16] British Library digital preservation strategy 2013-2016[EB/OL]. [2017-03-07]. http://www.bl.uk/aboutus/stratpolprog/collectioncare/discovermore/digitalpreservation/strategy/BL_DigitalPreservationStrategy_2013-16-external.pdf.

[17] 高娟 , 刘家真 . 中国大陆地区古籍数字化问题及对策 [J]. 中国图书馆学报 , 2013(4): 110-119.

[18] CADAL 项目资源入库总数 [EB/OL]. [2017-03-10]. http://www.cadal.cn/zydt/index1402.htm.

[19] BAND J.A guide for the perplexed part IV: the rejection of the Google Books settlement[EB/OL]. [2017-01-09]. http://www.districtdispatch.org/wp-content/uploads/2011/04/GuideIV-FINALV3.pdf.

[20] LCA releases statement on Google Books settlement rejection[EB/OL]. [2017-05-21]. http://old.arl.org/news/pr/gbs_statement24mar11~print.shtml.

[21] MARON N, PICKLE S. 数字投资评价 : ARL 各图书馆数字化特色馆藏的可持续性 [J]. 朱曼曼 , 译 . 图书情报

知识管理论坛，2017(5):406-413
DOI：10.13266/j.issn.2095-5472.2017.044

工作动态 , 2013(3): 15-16.
[22] Portico, Cornell University Library. Preservation of digitized books and other digital content held by cultural heritage organizations[EB/OL]. [2017-06-25]. http://www.portico.org/digital-preservation/wp-content/uploads/2010/01/NEH-IMLS-D-book-model.pdf.
[23] DOOLEY J, LUCE K. Taking our pulse: the OCLC research survey of special collections and archives[EB/OL]. [2017-06-25]. http://www.oclc.org/content/dam/research /publications/library/2010/ 2010-11.pdf?urlm=162945.
[24] 宋琳琳 , 李海涛 . 大型文献数字化项目元数据互操作调查与启示 [J]. 中国图书馆学报 , 2012(5): 27-38.
[25] Principles to guide vendor/publisher relations in large-scale digitization projects of special collections materials[EB/OL]. [2017-02-26]. http://www.arl.org/bm~doc/principles_large_scale_digitization.pdf.
[26] 莫莉·拉斐尔 . 数字化转变 : 电子书和电子内容对读者和图书馆的影响 [J]. 张钰羚 , 译 . 中国图书馆学报 , 2013(1): 31-35.
[27] 刘文琦 . 著作权法框架下大规模数字化复制的合理使用认定 [J]. 电子知识产权 , 2014(12): 69-73.
[28] 美法庭判决十年诉讼 : 谷歌将图书数字化合法 [EB/OL]. [2017-05-05]. http://tech.qq.com/a/20151017/008368.htm.
[29] MALPAS C. Cloud-sourcing research collections: managing print in the mass-digitized library environment[EB/OL]. [2017-01-02]. http://www.oclc.org/content/dam/research/publications/library/2011/2011-01.pdf?urlm=162949.
[30] History[EB/OL]. [2017-09-05]. https://dp.la/info/about/history/.
[31] Shared Print Program[EB/OL]. [2017-06-15]. https://www.hathitrust.org/shared_print_program.
[32] 欧洲数字图书馆和美国数字公共图书馆宣布合作 [J]. 现代图书情报技术 , 2012(3): 96.

作者贡献说明：
李咏梅：提出论文研究思路与框架，撰写论文；
刘　柳：负责数据收集；
刘　鸿：论文修订。

Research on the Mass Digitization of the Print Collection

Li Yongmei[1]　Liu Liu[1]　Liu Hong[2]
[1]Sichuan University Library, Chengdu 610065
[2]Library of Sichuan College of Architectural Technology, Deyang 618000

Abstract: [Purpose/significance] The mass digitization of the print collection is an inevitable action for digital libraries. It has enlarged the depth and breadth of library services, and has enormously promoted the development and progress of academic exchange and academic research. **[Method/process]** This paper analyzed the global mass digitization of the print collection in terms of the origin, environmental factors, development status and current problems. It took HathiTrust as a case study. **[Result/conclusion]** Building the principles, solving the copyright problems with fair use and establishing a multi-stage and multi-sided framework for cooperation are important measures to promote the full realization of the mass digitization.

Keywords: print collection　mass digitization　cloud library　HathiTrust　fair use